2T

УДК 519.863.001.63

В.А. Тазетдінов
Черкаський державний технологічний університет, м. Черкаси, Україна
valeriy_tazetdinov@rambler.ru

Нейромережна система аналізу і прогнозування процесів на ринку нерухомості

Розроблено інформаційно-аналітичну систему для аналізу і прогнозування стану ринку нерухомості. Виконано експериментальну перевірку ефективності функціонування інформаційно-аналітичної системи. Проведено верифікацію роботи розробленої системи з урахуванням зовнішніх факторів.

Вступ

Недостатнє будівництво нових об’єктів нерухомості, ріст промислового виробництва, відносно стабільна економічна ситуація, збільшення грошової маси і, як наслідок, зростання попиту на житло визначають зростання цін на нерухомість та спричиняють її дефіцит. В таких умовах органи державного управління, будівельні організації, які бажають продати чи купити житло, зацікавлені в одержанні інформаційно аналітичних послуг.

Аналіз процесів на ринку нерухомості (РН) свідчить про недостатню присутність методів математичного моделювання і аналітичних технологій в роботі ріелторів. До того ж сучасний стан надання аналітичних послуг не такий, як мав би бути, оскільки базується виключно на статистичному матеріалі без прив’язки до реальної ситуації та її особливостей.

Постановка задачі дослідження

Метою дослідження є розробка інформаційно-аналітичної системи (ІАС) оцінки об’єктів нерухомості (ОН), яка виконує ідентифікацію функції залежності ціни від внутрішніх факторів та її адаптацію до зовнішніх умов.

Необхідно розробити структуру, принципи функціонування інформаційно-аналітичної системи аналізу і прогнозування ринку нерухомості. Розроблювана система повинна проводити оцінку житла Z в залежності від його параметрів, в т.ч. і на макрорівні (з врахуванням особливостей міста, економічної ситуації в країні тощо), а також виконувати прогнозування тенденцій зміни цін на об’єкти нерухомості. Формалізацію цих задач виконано в [1-3] і буде використано при розробці ІАС.

ІАС потрібно наповнити таким змістом, який би сприяв грамотній постановці задачі, вибору оптимальної стратегії її розв’язку, відображав би стиль прийняття рішень у даній предметній області, а також саму модель цієї області. Таким чином, ця інтелектуальна система повинна мати властивості адаптації та самоорганізації. Її орієнтованість на конкретного споживача чи коло споживачів, які мають власні уявлення з інтерпретації, оцінки, обробки та аналізу інформації, дозволить значно підвищити якість рішень, а сама інформаційно-аналітична система стане інтелектуальним партнером людини при розв’язуванні задачі [4], [5]. В ній будуть інтегровані знання математика, програміста та системного аналітика з проблем ринку нерухомості.

Розроблена ІАС повинна надавати аналітичні послуги суб'єктам РН і класифікувати ОН за заданими еталонами; визначати тенденції руху цін на однотипні ОН; визначати пріоритетні фактори і міру їх впливу на ціну ОН; знаходити еталон в класі, якому належить ОН; розраховувати діапазон, в якому може бути продано ОН (песимістична ціна – оптимістична ціна); визначати тенденції зміни однотипних ОН за різними факторами та розв'язувати супутні задачі.

Запропонована для використання в ІАС нейромережна технологія є ефективною при використанні у клієнт-серверній архітектурі, при знаходженні бази даних на сервері, а програмно-аналітичного забезпечення на робочій станції. Її використання значно зменшує час оцінки об'єкта та інформаційного супроводу його продажу.

Структура та особливості реалізації інформаційно-аналітичної системи REMA

Широке впровадження і визначення переваг технологій data mining [6], [7], knowledge discovery in databases (KDD) [8], OLAP дає всі підстави для застосування штучних нейронних мереж. Особливості задач аналізу РН вимагають застосування елементів всіх перерахованих технологій. Так, необхідність знаходження нових знань серед великої кількості даних визначає потребу у методах data mining, значні розміри інформаційних баз вказують на переваги KDD, багатовимірність простору даних свідчить про ефективність застосування аналітичних методів OLAP. Попередній аналіз показав, що найбільш придатними для базових задач аналізу РН є прямозв'язна мережа зустрічного поширення (МЗП) і мережа із алгоритмом оберненого поширення помилки (АОПП) за умови застосування в них стохастичного (больцманівського) методу навчання із використанням як активаційних функцій щільностей розподілів Коші, Лапласа і Гауса. Так, відомо, що на вхід МЗП подаються значення векторів X і Z, для нашої задачі $Z = (Z_{\min}, Z_{реал}, Z_{\max})$. Після процесу навчання на вхід мережі достатньо подавати на вибір або лише елементи вектора X, або лише вектора Z, або лише значення одного фактора із масивів X і Z. Це дає можливість ідентифікації залежності ціни від одного фактора при незмінних інших, а визначення елементів вектора Z вимагає ще і встановлення рівня оптимістичної, песимістичної та реальної ціни на ОН.

Зауважимо, що МЗП реалізує швидкий алгоритм навчання, але точність результатів є достатньою лише для попереднього аналізу. Значно точнішими є результати функціонування нейронної мережі з АОПП. На вхід мережі подається лише вектор X із n компонент, вихід має три складові. В режимі використання мережі на вхід досить подавати значення одного фактора, який нас цікавить, а замість інших – значення їх середніх величин.

Технологія аналізу та прогнозування РН реалізована в ІАС REMA (Real Estate Management Analysis). Запропонована технологія аналізу РН дозволяє здійснювати аналіз тенденцій, які складаються на ринку, прогнозувати динаміку зміни ціни ОН. Відштовхуючись від середньостатистичних оцінок ОН, розроблено метод їх адаптації до особливостей міста, регіону, столиці та коригування з врахуванням макроекономічної ситуації, що склалася в країні. Встановлено, що оптимальним для цієї процедури є використання лінгвістичних змінних. Розроблено технологію їх формалізації та здійснено порівняльний аналіз методів ідентифікації, який засвідчив оптимальність використання штучних нейронних мереж із стохастичним алгоритмом навчання.

Використання вищенаведених компонент в ІАС REMA дозволяє здійснювати підтримку прийняття рішень за рахунок розв'язання задач аналізу, синтезу та вибору оптимальних розв'язків при відомих альтернативах. Для цього в REMA згідно з принципом свободи вибору існує можливість вибору моделі із певного класу, здійснення її структурної та параметричної оптимізації та виконання безпосередньо моделювання.

ІАС REMA є засобом створення, ведення, коригування інформаційного банку (ІБ), який містить інформацію про ОН, особливості законодавчої бази, основні статистичні константи та інші дані, а також призначена для аналізу РН та прогнозування тенденцій.

На базі розробленої технології дослідження РН ІАС REMA дозволяє реалізувати таку стратегію його аналізу та прогнозування:

1. Відповідно до цілей дослідження здійснювати синтез моделі, вибір інформаційного базису, констант та методів дослідження.

2. При оцінці одержаних результатів згідно з критеріальними функціями здійснювати аналіз ефективності алгоритмів, які використовувались, і при необхідності здійснювати корекцію моделі, настройку коефіцієнтів та вибирати альтернативні моделі та методи аналізу.

3. При функціонуванні ІАС здійснювати інтерактивну взаємодію із користувачем, необхідність якої може бути викликана особливістю моделювання нейронної мережі (НМ), можливістю її паралічу та іншими причинами.

4. Згідно з принципом інформаційної єдності здійснювати наскрізне супроводження процесу обробки даних з протоколюванням та видачею головних проміжних результатів у файл.

5. Здійснювати ідентифікацію законів зміни значень факторів, аналіз та прогнозування.

6. Зменшувати апріорну ентропію вихідних даних за рахунок попередньої обробки даних, в т.ч. і нейромережними методами.

7. Здійснювати верифікацію одержаних результатів з використанням альтернативних методів дослідження.

8. Забезпечувати можливість функціонування відкритої архітектури ІАС REMA, що пов'язано із постійним розширенням спектра задач на РН.

Функціональне навантаження, яке виконує ІАС REMA, розподілене на декілька фрагментів, які визначаються суб'єктом, що працює з системою (рис. 1). Інформаційну основу складає ІБ, який є трикомпонентним і включає в себе банк пропозицій ($ІБ_1$), банк попиту ($ІБ_2$) та банк, який містить довідкову інформацію ($ІБ_3$). Всі компоненти знаходяться на сервері та є інформаційно незалежними один від іншого. Введення даних та їх коригування в $ІБ_1$ та $ІБ_2$ здійснюється оператором. В разі необхідності оператор може отримати потрібну інформацію із таблиць інформаційного банку за SQL-запитами. Формування $ІБ_3$, який містить закони, що визначають функціонування РН, статистичні таблиці, інші константи, виконує спеціаліст-аналітик.

Дані інформаційного банку використовують фахівці на робочих станціях. ІАС інтегрує в собі тактику дослідження, яка визначається виконанням перетворення даних у знання, шляхом здійснення аналітичних розрахунків з використанням нейромережної парадигми та допоміжних процедур і відображенням їх у простір консультативних послуг.

Враховуючи те, що в $ІБ_1$ знаходяться дані про ОН, включно із реалізованими, за допомогою НМ із стохастичним алгоритмом навчання, виконаємо ідентифікацію залежностей

$$Z_б = F_1(X) \text{ і } Z_p = F_2(X), \qquad (1)$$

де F_1 – вартість нереалізованих, F_2 – реалізованих ОН, $Z_б$ та Z_p – бажана та реальна ціна.

Знання функцій F_1 та F_2 дозволяє визначити бажану для продавців ціну ОН та реальну ціну, тобто ту, за яку він може бути проданий у найкоротший термін. Залежності (1) також дозволяють визначити чутливість ціни ОН до зміни значень факторів. Практичне значення цього знання полягає у можливості зближення реальної та бажаної ціни, а також визначення найбільш значущих факторів, що дозволить її обґрунтовано варіювати.

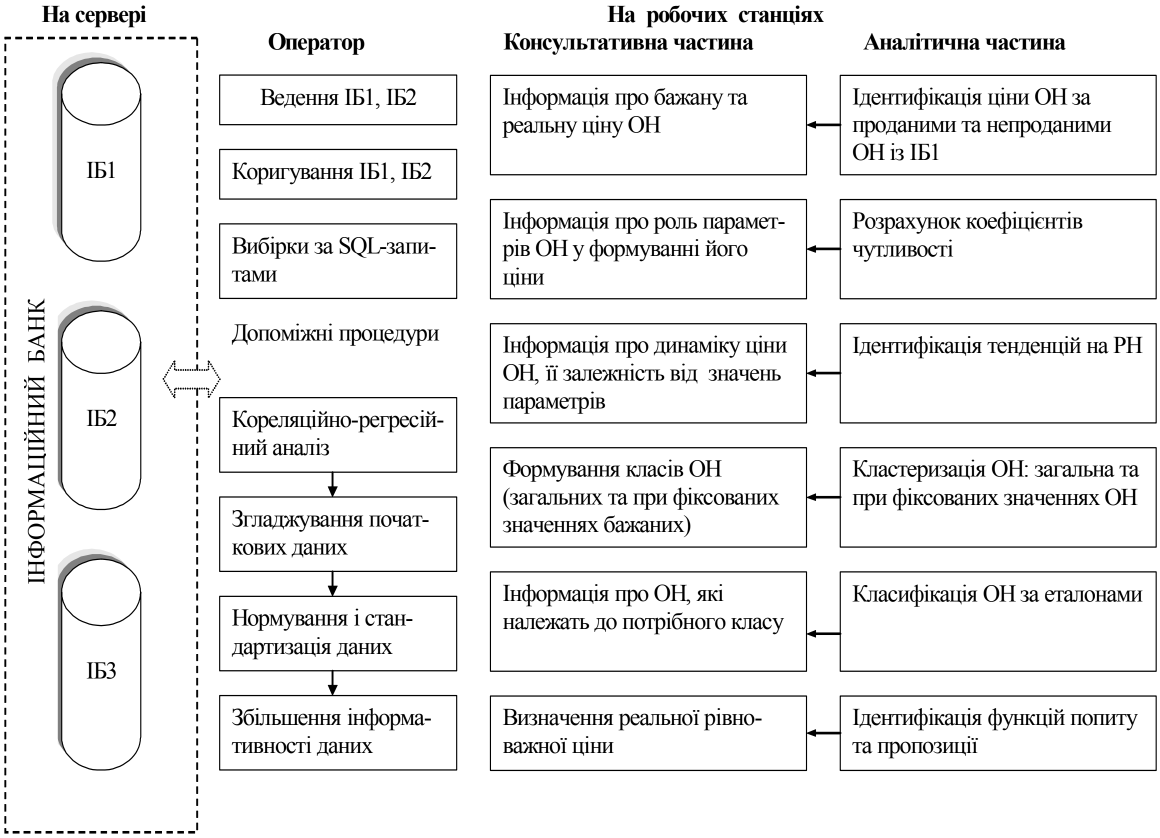


Рисунок 1 – Структурна схема інформаційно-аналітичної системи REMA

Узагальнення результатів експерименту із урахуванням впливу зовнішніх факторів

Для ідентифікації та дослідження залежності ціни ОН від внутрішніх та зовнішніх факторів необхідно також знати суб'єктивні оцінки місцевих, регіональних та макроекономічних факторів. Інформаційною базою розв'язання такої задачі визначимо пропозиції на РН нерухомості у м. Черкасах.

Важливою складовою ідентифікованої залежності є формалізовані значення суб'єктивних факторів. Оскільки такі значення є стохастичними, то одержати їх функціональні залежності від часу неможливо. Саме тому функції належності будемо вважати дискретними, а їх побудову відносити до компетентності експертів. В IAC REMA передбачено поряд із введенням інформації про ОН (значень внутрішніх параметрів) також задання значень суб'єктивних факторів [9]. При цьому враховано, що певна інформація записується в $ІБ_1$ автоматично, згідно із районом розміщення ОН. Інші значення записуються автоматично протягом певного проміжку часу.

Значення функцій належності визначаємо шляхом експертного опитування на базі карти районів міста та формалізованих суб'єктивних факторів (табл. 1). Так, для районів міста одержимо:

$$\mu(L_1^p) = \{\frac{віддалений}{0,2}; \frac{центральний}{1}; \frac{близький\ до\ центрального}{0,8}\},$$

$$\mu(L_2^p) = \{\frac{дуже\ погане}{0,2}; \frac{погане}{0,4}; \frac{не\ досить\ добре}{0,6}; \frac{добре}{0,8}; \frac{дуже\ добре}{1}\},$$

$$\mu(L_3^p) = \{\frac{дуже\ забруднений}{0,2}; \frac{забруднений}{0,4}; \frac{задовільний\ стан}{0,6}; \frac{екол.\ чистий}{1}\},$$

$$\mu(L_4^p) = \{\frac{немає}{0,2}; \frac{не\ дуже\ розвинена}{0,4}; \frac{нормальна}{0,6}; \frac{добра}{0,8}\}.$$

Таблиця 1 – Значення функцій належності для факторів, які характеризують район

Фактор / Район	10	20	30	40	50	60	70	80	90
Відстань від центру	1	0,8	0,8	0,2	0,8	0,8	0,8	0,2	0,8
Транспорт. забезп.	1	0,6	0,4	0,4	0,9	0,6	0,4	0,4	0,8
Екологія	0,2	0,6	0,4	0,6	0,6	0,6	0,4	0,2	0,4
Інфраструктура	0,8	0,6	0,6	0,6	0,8	0,6	0,6	0,2	0,4

Значення функцій належності для факторів, які характеризують стан РН у місті, визначаються часом і у загальному випадку можуть бути такими:

$$\mu(L_1^м) = \{\frac{низький}{0,8}; \frac{середній}{0,6}; \frac{високий}{0,3}; \frac{дуже\ високий}{0,1}\},$$

$$\mu(L_2^м) = \{\frac{низьке}{0,2}; \frac{рівноважне}{0,5}; \frac{високе}{0,8}\},$$

$$\mu(L_3^м) = \{\frac{1:3}{0,8}; \frac{1:2}{0,6}; \frac{1:1}{0,5}; \frac{\geq 2:1}{0,1}\},$$

$$\mu(L_4^м) = \{\frac{пасивний\ ринок}{0,2}; \frac{помірний}{0,4}; \frac{стабільний}{0,8}; \frac{активний}{0,7}\}.$$

Для сучасної ситуації на РН: $\mu(L_1^м) = 0,3;\ \mu(L_2^м) = 0,2;\ \mu(L_3^м) = 0,5;\ \mu(L_4^м) = 0,7$.

Регіональні особливості можуть мати такі фактори та значення функцій належності:

$\mu(L_1^в) = \{$*переважають сировинні галузі*$/0{,}2;$ *орієнтований на виробництво сільськогосподарської продукції* $/0{,}4;$ *має розвинений агропромисловий комплекс* $/ 0{,}6;$ *промисловий регіон*$/0{,}7;$ *переважають наукомісткі виробництва*$/0{,}9\};$

$$\mu(L_2^в) = \{\frac{найнижчі}{0{,}2}; \frac{низькі}{0{,}3}; \frac{нижче\ середніх}{0{,}4}; \frac{середні}{0{,}6}; \frac{вище\ середніх}{0{,}9}\},$$

$$\mu(L_3^в) = \{\frac{низька}{0{,}3}; \frac{середня}{0{,}5}; \frac{висока}{0{,}7}\},$$

$$\mu(L_4^в) = \{\frac{забруднений}{0{,}2}; \frac{задовільний}{0{,}4}; \frac{добрий}{0{,}6}; \frac{екологічно\ чистий}{0{,}9}\}.$$

Черкаський регіон характеризуємо такими значеннями:

$\mu(L_1^в) = 0{,}4;\ \mu(L_2^в) = 0{,}6;\ \mu(L_3^в) = 0{,}5;\ \mu(L_4^в) = 0{,}4.$

Можливим варіантам макроекономічної ситуації в країні відповідають функції належності:

$$\mu(L_1^a) = \{\frac{<1\%}{0{,}9}; \frac{2-5\%}{0{,}6}; \frac{5-15\%}{0{,}4}; \frac{>15\%}{0{,}1}\},$$

$$\mu(L_2^a) = \{\frac{<1\%}{0{,}9}; \frac{2-5\%}{0{,}7}; \frac{5-15\%}{0{,}4}; \frac{>15\%}{0{,}2}\},$$

$$\mu(L_3^a) = \{\frac{<1\%}{0{,}9}; \frac{2-5\%}{0{,}6}; \frac{5-15\%}{0{,}4}; \frac{>15\%}{0{,}1}\},$$

$$\mu(L_4^a) = \{\frac{низькі}{0{,}9}; \frac{помірні}{0{,}6}; \frac{середні}{0{,}5}; \frac{високі}{0{,}2}\}.$$

Оскільки наповнення інформаційного банку за досить активного РН постійно змінювалось, то інформація про особливості міста, регіону та макроекономічну ситуацію в країні залишалась постійною і не змінювалась. Тому активну участь в аналізі цін на РН брали лише фактори, які визначали районні особливості місцерозташування ОН. Було також враховано, що коефіцієнт порівняння районних цін із столичними становив 1,9.

В результаті використання НМ із модифікованим алгоритмом стохастичної релаксації (МАСР) для ідентифікації функції оцінки ОН значення цільової функції (середнього абсолютного відхилення табличних значень від розрахованих) покращувалось на 2 – 4,5 % в залежності від експерименту, що свідчить на користь врахування зовнішніх факторів.

Експерименти проводились у системі, конфігурація якої була такою: сервер з процесором Pentium IV-2,4, 512 Мб ОЗУ та операційною системою Windows 2000 Server; робоча станція з процесором AMD Athlon-2600, 256 Мб ОЗУ та операційною системою Windows XP. Вихідними даними була інформація про 500 ОН з $ІБ_1$, який знаходився на сервері. На робочій станції функціонувала ІАС REMA. Час навчання НМ становив 16,2 хв. при заданій точності 0,25 на нормованій навчальній вибірці та співвідношенні навчальної вибірки до контрольної 60:40. При пропорції 70:30 час навчання збільшувався на 14 – 20 % у порівнянні з попереднім варіантом, якщо ж співвідношення було 80:20, то час збільшувався ще на 10 – 16 %. Точність ідентифікації, яка перевірялась за середньою абсолютною похибкою, у залежності від

варіанта співвідношення кількості ОН у навчальній та контрольній вибірці, змінювалась на 0,1 – 0,65 %, що свідчило про достатньо велику кількість ОН для навчання та ефективність застосування НМ із МАСР.

В умовах повнорозмірного інформаційного банку та великої кількості факторів, які необхідно врахувати, переваги алгоритмів із стохастичними принципами навчання є беззаперечними, оскільки сучасний рівень розвитку технічних засобів дозволяє розв'язувати більшість практичних задач за допомогою НМ за прийнятний час, а головним критерієм оцінки якості одержаних розв'язків є їх точність. Швидкість збіжності методу МАСР з використанням різних розподілів є більшою на 6 – 24 % за інші методи стохастичного навчання.

Проведені експерименти на реальних даних (ІБ містив 840 записів) з використанням МАСР вказують на наступне. По-перше, мережа навчалась в середньому на 2 – 4 % швидше, ніж при використанні класичного АОПП, і на 22 % швидше, ніж при використанні навчання за алгоритмом машини Больцмана.

Висновки

Сучасний стан ринку нерухомості та майже повна відсутність методів його аналізу та прогнозування свідчать про необхідність розвитку інформаційно-аналітичних систем. Ці системи функціонуватимуть на єдиній інформаційній базі об'єктів нерухомості, дозволяючи державним органам управління прогнозувати тенденції та спрямування розвитку РН, будівельним організаціям планувати структуру виробництва, агентствам нерухомості вивчати перспективи РН, а також продавцям і покупцям РН повністю задовольняти інформаційні потреби.

Тенденції на РН значно впливають на його функціонування, оскільки визначають як майбутній, так і теперішній попит та пропозицію на ОН, що дозволяє визначитись із часом його купівлі чи продажу. Знаходження залежності ціни ОН від значення певного фактора в часі допомагає встановити чутливість динаміки ціни до зміни значення фактора в часі.

Однією із головних функцій розробленої ІАС є прогнозування. Після того як шукані залежності ідентифіковані, можна задавати запити у формі «а якщо А, то що...?». Передбачення таким чином можливих сценаріїв розвитку ринку нерухомості спрощує процеси прийняття рішень та підвищує їх достовірність.

Література

1. Тазетдінов В.А. Самоорганізація інформаційної бази ринку нерухомості на основі нейромережевих технологій // АСУ и приборы автоматики. – 2004. – Вып. 127. – С. 41-47.
2. Тазетдинов В.А. Технология нейросетевого прогнозирования рынка недвижимости // Искусственный интеллект. – 2004. – № 3. – С. 593-597.
3. Тазетдінов В.А. Адаптивна технологія нечіткого аналізу динаміки ринку нерухомості // Радіоелектроніка і інформатика. – 2004. – № 3 (28). – С. 126-129.
4. Снитюк В.Є., Говорухін С.О. Технологія data mining і засоби її реалізації // Вісник ЧДТУ. – 2002. – № 3. – С. 80-84.
5. Астанин С.В. Особенности анализа и моделирования систем гибридного интеллекта // УСИМ. – 2001. – № 1. – С. 16-23.
6. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах // Открытые системы. – 1997. – № 4. – С. 41-44.

7. Шапот М. Интеллектуальный анализ данных в системах поддержки принятия решений // Открытые системы. – 1998. – № 1. – С. 30-35.
8. Глибовець М.М., Олецький О.В. Штучний інтелект. – К.: КМ «Академія», 2002. – 366 с.
9. Тазетдінов В.А. Інтелектуалізація алгоритму функціонування інформаційно-аналітичного забезпечення ринку нерухомості // Збірник доп. Міжн. наук.-практ. конф. «Единое информационное пространство». – Дніпропетровськ. – 2003. – С. 158-159.

В.А. Тазетдинов
Нейросетевая система анализа и прогнозирования процессов на рынке недвижимости
Разработана информационно-аналитическая система для анализа и прогнозирования состояния рынка недвижимости. Выполнена экспериментальная проверка эффективности функционирования информационно-аналитической системы. Проведена верификация работы разработанной системы с учетом внешних факторов.

V.A. Tazetdinov
Real Estate Market Prediction and Analysis Neural Nets System
The informational analytical system for analysis and prediction of real estate market’s state has been developed. The experimental validation of effective operation of informational analytical system has been carried out. The verification of working of the developed system taking into account outside factors has been accomplished.

Стаття надійшла до редакції 03.04.2008.